This handout includes space for every question that requires a written response. Please feel free to use it to handwrite your solutions (legibly, please). If you choose to typeset your solutions, the `README.md` for this assignment includes instructions to regenerate this handout with your typeset LaTeX solutions.

1.

Given a unit vector u, we can decompose each x vector into 2 components:

- the projection of the x vectors onto a vector u, call it p_{xu},
- the residual, $r_{xu} = x - p_{xu}$

Note that r_{xu} is orthogonal to p_{xu}, i.e. the dot product is zero, $r_{xu} * p_{xu} = 0$ Now, the total variance of the x vectors is

$$\begin{aligned} var(x) &= x' * x \\ &= (p_{xu} + r_{xu})' * (p_{xu} + r_{xu}) \\ &= (p_{xu})' * p_{xu} + (r_{xu})' * r_{xu} \text{ (cross terms drop due to orthogonality)} \end{aligned}$$

So by maximizing one of the terms in the last equation we also minimize the other. Since PCA maximizes the first it also minimizes the second.

2.a

$$
\begin{aligned}
\ell(\theta^{(t+1)}) &= \alpha\ell_{\textsf{sup}}(\theta^{(t+1)}) + \ell_{\textsf{unsup}}(\theta^{(t+1)}) && \text{Definition} \\
&\geq \alpha\ell_{\textsf{sup}}(\theta^{(t+1)}) + \sum_{i=1}^{n}\sum_{z^{(i)}} Q_i^{(t)}(z^{(i)}) \log \frac{p(x^{(i)}, z^{(i)}; \theta^{(t+1)})}{Q_i^{(t)}(z^{(i)})} && \text{Jensen's inequality} \\
&\geq
\end{aligned}
$$

2.b

2.c

List the parameters which need to be re-estimated in the M-step:

In order to simplify derivation, it is useful to denote

$$w_j^{(i)} = Q_i^{(t)}(z^{(i)} = j),$$

and

$$\tilde{w}_j^{(i)} = \begin{cases} \alpha & \tilde{z}^{(i)} = j \\ 0 & \text{otherwise.} \end{cases}$$

We further denote $S = \Sigma^{-1}$, and note that because of chain rule of calculus, $\nabla_S \ell = 0 \Rightarrow \nabla_\Sigma \ell = 0$. So we choose to rewrite the M-step in terms of S and maximize it w.r.t S, and re-express the resulting solution back in terms of Σ.

Based on this, the M-step becomes:

$$\begin{aligned}
\phi^{(t+1)}, \mu^{(t+1)}, S^{(t+1)} &= \arg\max_{\phi,\mu,S} \sum_{i=1}^{n} \sum_{j=1}^{k} Q_i^{(t)}(z^{(i)}) \log \frac{p(x^{(i)}, z^{(i)}; \phi, \mu, S)}{Q_i^{(t)}(z^{(i)})} + \alpha \sum_{i=1}^{\tilde{n}} \log p(\tilde{x^{(i)}}, \tilde{z^{(i)}}; \phi, \mu, S) \\
&=
\end{aligned}$$

Now, calculate the update steps by maximizing the expression within the argmax for each parameter (We will do the first for you).

ϕ_j: We construct the Lagrangian including the constraint that $\sum_{j=1}^k \phi_j = 1$, and absorbing all irrelevant terms into constant C:

$$\begin{aligned}
\mathcal{L}(\phi, \beta) &= C + \sum_{i=1}^{n} \sum_{j=1}^{k} w_j^{(i)} \log \phi_j + \sum_{i=1}^{\tilde{n}} \sum_{j=1}^{k} \tilde{w}_j^{(i)} \log \phi_j + \beta \left(\sum_{j=1}^{k} \phi_j - 1 \right) \\
\nabla_{\phi_j} \mathcal{L}(\phi, \beta) &= \sum_{i=1}^{n} w_j^{(i)} \frac{1}{\phi_j} + \sum_{i=1}^{\tilde{n}} \tilde{w}_j^{(i)} \frac{1}{\phi_j} + \beta = 0 \\
\Rightarrow \phi_j &= \frac{\sum_{i=1}^{n} w_j^{(i)} + \sum_{i=1}^{\tilde{n}} \tilde{w}_j^{(i)}}{-\beta} \\
\nabla_\beta \mathcal{L}(\phi, \beta) &= \sum_{j=1}^{k} \phi_j - 1 = 0 \\
\Rightarrow \sum_{j=1}^{k} \frac{\sum_{i=1}^{n} w_j^{(i)} + \sum_{i=1}^{\tilde{n}} \tilde{w}_j^{(i)}}{-\beta} &= 1 \\
\Rightarrow -\beta &= \sum_{j=1}^{k} \left(\sum_{i=1}^{n} w_j^{(i)} + \sum_{i=1}^{\tilde{n}} \tilde{w}_j^{(i)} \right) \\
\Rightarrow \phi_j^{(t+1)} &= \frac{\sum_{i=1}^{n} w_j^{(i)} + \sum_{i=1}^{\tilde{n}} \tilde{w}_j^{(i)}}{\sum_{j=1}^{k} \left(\sum_{i=1}^{n} w_j^{(i)} + \sum_{i=1}^{\tilde{n}} \tilde{w}_j^{(i)} \right)} \\
&= \frac{\sum_{i=1}^{n} w_j^{(i)} + \sum_{i=1}^{\tilde{n}} \tilde{w}_j^{(i)}}{n + \alpha \tilde{n}}
\end{aligned}$$

μ_j: Next, derive the update for μ_j. Do this by maximizing the expression with the argmax above with respect to μ_j.

First, calculate the gradient with respect to μ_j:

$$\nabla_{\mu_j} =$$

Next, set the gradient to zero and solve for μ_j:

$$0 =$$

$\mathbf{\Sigma}_j$: Finally, derive the update for Σ_j via S_j. Again, Do this by maximizing the expression with the argmax above with respect to S_j.

.

First, calculate the gradient with respect to S_j:

$$\nabla_{S_j} =$$

Next, set the gradient to zero and solve for S_j:

$$0 =$$

This results in the final set of update expressions:

$$\begin{aligned}\phi_j &:= \\ \mu_j &:= \\ \Sigma_j &:=\end{aligned}$$

2.f

3.a

3.b